



## PERSPECTIVE



## Cross-ancestry genomic research: time to close the gap

Elizabeth G. Atkinson<sup>1</sup>, Sevim B. Bianchi<sup>2</sup>, Gordon Y. Ye<sup>1,2</sup>, José Jaime Martínez-Magaña<sup>3,4</sup>, Grace E. Tietz<sup>1</sup>,  
Janitza L. Montalvo-Ortiz<sup>1,3,4,5</sup>, Paola Giusti-Rodriguez<sup>6</sup>, Abraham A. Palmer<sup>1,2,7</sup> and Sandra Sanchez-Roige<sup>1,2,8</sup>✉

© The Author(s), under exclusive licence to American College of Neuropsychopharmacology 2022

Neuropsychopharmacology (2022) 47:1737–1738; <https://doi.org/10.1038/s41386-022-01365-7>

Genome-wide association studies (GWAS) have revolutionized our ability to understand the genetic underpinnings of biomedical traits; however, their extreme Eurocentric bias has exacerbated health inequities. In this perspective, we highlight recent efforts to address the imbalance of ancestral representation in genomics research, including the formation of large collaborative efforts and the development of novel statistical methods to improve translation of genomic insights across ancestries. Using more ancestrally diverse GWAS samples will improve our understanding of the genetic architecture of complex diseases, not only for the understudied populations, but for individuals of all ancestral backgrounds.

GWAS have yielded a wealth of clues about the molecular basis of many common human diseases [1]. Firstly, GWAS have repeatedly identified associations with genes that are the targets of existing and highly effective pharmacologic agents, such as *PCSK9* for cardiovascular health [2], and *DRD2* for schizophrenia [3], and may allow for the identification of additional novel biological pathways. Secondly, GWAS have revealed that virtually all traits are influenced by many variants, each with small effect sizes, distributed throughout the genome. Using GWAS results, we can combine the risk conferred by these multiple variants into a single genetic liability score (i.e., polygenic score (PGS)), which may assist in risk prediction and disease stratification [4], potentially contributing to precision medicine. Despite these advances, there is a significant problem with GWAS: most current well-powered GWAS are performed on samples with a drastic overrepresentation of individuals of European ancestry. This is problematic because genomic results often do not fully transfer across ancestries [5]. Although it is presumed that people of all ancestries share the same underlying biological disease mechanisms and causal variants, some variants may be more frequent or more correlated in different genetic backgrounds. As a result, variants may be statistically correlated with a trait in one ancestral group but not another. Similarly, while PGS might perform well within a specific ancestry, their accuracy decreases when applied to others [6–8]. This lack of diversity in gene discovery cohort composition alongside a paucity of methods designed for diverse populations has resulted in reduced transferability of findings across ancestries, which may exacerbate existing health disparities and stigmatization [9, 10].

One way to address this problem is by generating GWAS datasets in reference panels based on individuals from diverse ancestries. Current efforts have been propelled by both academic and direct-to-consumer genetic companies (e.g., *All of Us*, Million Veterans Program, China Kadoorie Biobank, Biobank Japan, TOPMed, 23andMe), alongside large-scale data aggregation largely led by consortia (e.g., Latin American Genomics Consortium, H3Africa, PAGE, Qatar Biobank, GenomeAsia 100k, gnomAD). To support these massive efforts, it is of critical importance to facilitate funding mechanisms. Of equal importance are targeted outreach and education efforts (e.g., workshops, community engagement, and development and distribution of informational materials) that build trust in genomic research among minority populations, ensuring they benefit maximally from the research.

However, collecting the vast amounts of data needed to diversify our datasets is an immense undertaking. Along with the growing awareness that genomic studies need to include more diverse populations, there is also a push to improve methods for studying data from these populations. Some of the under-represented populations that may fill this gap are genetically heterogeneous and contain genetic components from multiple continental ancestries, also known as “admixed”. For example, Latin American and African American individuals are typically admixed between two or three different continental ancestries. Admixed populations have generally been excluded from GWAS due to the difficulty of effectively accounting for their complex genomic structure. One promising strategy to account for this structure is the use of local ancestry (i.e., the particular ancestry of each genomic segment of an individual). Early association efforts in admixed cohorts utilized local ancestry via admixture mapping and novel tools are being developed that build local ancestry into GWAS [11], including *Tractor* [12]. Other recent works have developed ancestry-aware methods that only require summary statistics (e.g., Multi-Ancestry Meta-Analysis [13]). Applying these multi-ancestry genomics methods to combine more samples will consequently increase power to detect genetic factors for complex traits shared across ancestries, and help localize signals closer to causal variants (e.g. [14]). In addition to GWAS, other polygenic methods are being actively developed to better estimate heritability [15], generate cross-ancestry genetic correlations [16], and improve the transferability of PGS across

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. <sup>2</sup>Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. <sup>3</sup>Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. <sup>4</sup>VA CT Healthcare Center, West Haven, CT, USA. <sup>5</sup>US Department of Veterans Affairs National Center of Posttraumatic Stress Disorder, Clinical Neurosciences Division, West Haven, CT, USA. <sup>6</sup>Department of Psychiatry, University of Florida College of Medicine, Gainesville, FL, USA. <sup>7</sup>Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. <sup>8</sup>Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ✉email: [sanchezroige@ucsd.edu](mailto:sanchezroige@ucsd.edu)

populations [17–20]. Along with other sources of omic data (many of which are also currently Eurocentric), novel methods leverage cross-population prediction at the level of gene transcript [21], cell-type specific regulatory annotations [22], and even gene network analyses.

Despite much progress, many current methods do not account for the complex sociocultural experiences of individuals that may impact health outcomes or disease prevalence [23]. When ancestry-specific results arise, we must be cognizant that these may reflect differences in case ascertainment as well as environmental exposures, societal factors, and demographics (e.g., socioeconomic status, diet) that may be confounded by ancestry. Special attention is needed to ensure that the phenotypes and ancestry categorization in understudied groups are well-defined.

Beyond these efforts to generate more diverse data and enhance methods for its analysis, other challenges remain to increase equity in genomics research [5, 9, 10, 24]. To work toward closing the diversity gap, it is important that there is sufficient and sustained support for efforts to increase the diversity of scientists through training programs and capacity building (e.g., <https://gingerprogram.org>), and the promotion and recognition of local researchers. Current and future consortium efforts should maintain equitable and ethical partnerships with low- and middle-income countries (LMIC), ensuring that they represent full partners and not data harvesters. There must also be an increased effort to revise current publication and grantmaking policies to ensure that they do not disadvantage researchers from underrepresented communities. The success of these initiatives will require support from funding agencies and scientific journals, for example, by considering studies of all cohort sizes, encouraging replication of findings in distinct ancestries, offering fee waivers for publication and open access to journals, particularly in LMICs, and flexible data sharing policies [25]. These efforts will ensure that the benefits of genomics research are shared across populations, striving toward global health equity.

## REFERENCES

- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet.* 2017;101:5–22.
- Shapiro MD, Tavori H, Fazio S. PCSK9: from basic science discoveries to clinical trials. *Circ Res.* 2018;122:1420–38.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511:421–7.
- Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 2007;17:1520–8.
- Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 2019;179:589–603.
- Duncan L, Shen H, Gelaye B, Meijssen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun.* 2019;10:3328.
- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat Genet.* 2019;51:584–91.
- Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017;100:635–49.
- Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med.* 2022;28:243–50.
- Palk AC, Dalvie S, de Vries J, Martin AR, Stein DJ. Potential use of clinical polygenic risk scores in psychiatry—ethical implications and communicating high polygenic risk. *Philos Ethics Humanit Med.* 2019;1:4.
- Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. *Nat Rev Genet.* 2011;12:523–8.
- Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, et al. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat Genet.* 2021;53:195–204.
- Turley P, Martin AR, Goldman G, Li H, Walters RK, Jala JB, et al. Multi-ancestry meta-analysis yields novel genetic discoveries and ancestry-specific associations. *bioRxiv.* 2021.
- Zhong Y, De T, Alarcon C, Park CS, Lec B, Perera MA. Discovery of novel hepatocyte eQTLs in African Americans. *PLoS Genet.* 2020;16:e1008662.
- Luo Y, Li X, Wang X, Gazal S, Mercader JM, 23andMe Research Team, et al. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Hum Mol Genet.* 2021;30:1521–34.
- Zhang J, Schumacher FR. Evaluating the estimation of genetic correlation and heritability using summary statistics. *Mol Genet Genomics.* 2021;296:1221–34.
- Hoggart C, Choi S, Preuss M, O'Reilly P. BridgePRS: a powerful trans-ancestry polygenic risk score method. *Europe PMC.* 2022.
- Ruan Y, Lin Y-F, Feng Y-CA, Chen C-Y, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *medRxiv.* 2021.
- Weissbrod O, Kanai M, Shi H, Gazal S, Peyrot WJ, Khera AV, et al. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat Genet.* 2022;54:450–8.
- Zhang H, Zhan J, Jin J, Zhang J, Ahearn TU, Yu Z, et al. Novel methods for multi-ancestry polygenic prediction and their evaluations in 3.7 million individuals of diverse ancestry. *bioRxiv.* 2022.
- Liang Y, Pividori M, Manichaikul A, Palmer AA, Cox NJ, Wheeler HE, et al. Polygenic transcriptome risk scores (PTRS) can improve portability of polygenic risk scores across ancestries. *Genome Biol.* 2022;23:23.
- Amariuta T, Ishigaki K, Sugishita H, Ohta T, Koido M, Dey KK, et al. Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat Genet.* 2020;52:1346–54.
- McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, et al. Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *Am J Epidemiol.* 2017;186:753–61.
- Atutornu J, Milne R, Costa A, Patch C, Middleton A. Towards equitable and trustworthy genomics research. *EBioMedicine.* 2022;76:103879.
- Atkinson E, Choquet H, Khor CC, Wonkam A. Improving equity in human genomics research. *Commun Biol.* 2022;5:281.

## AUTHOR CONTRIBUTIONS

SSR conceived the idea. SSR and EGA wrote the first draft of the manuscript. All authors edited and approved the final version of the manuscript.

## FUNDING

This work is supported by the California Tobacco-Related Disease Research Program (28IR-0070 to AAP and T29KT0526 to SSR and SBB), NIDA (DP1DA054394 to SSR), by the Department of Veterans Affairs via NIDA (R21DA050160 and 1K2CX002095-01A1 to JLMO), the National Institutes of Mental Health (K01 MH121659 to EGA), the National Institute of General Medical Sciences (GM139534-01 to GET), and the U.S. Department of Veterans Affairs National Center for Posttraumatic Stress Disorder. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Sandra Sanchez-Roige.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.